AD_____

Award Number: W81XWH-04-1-0271

TITLE: Fatty Acid Synthesis Gene Variants and Breast Cancer Risk: A Study within the European Prospective Investigation into Cancer and Nutrition

PRINCIPAL INVESTIGATOR: Rudolf J. Kaaks, Ph.D.

CONTRACTING ORGANIZATION: International Agency for Research on Cancer
Lyon, France 69008

REPORT DATE: February 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 2005	Annual (30 Jan 2004 – 30 Jan 2005)

4. TITLE AND SUBTITLE
Fatty Acid Synthesis Gene Variants and Breast Cancer Risk: A Study within the European Prospective Investigation into Cancer and Nutrition

5. FUNDING NUMBERS
W81XWH-04-1-0271

6. AUTHOR(S)

Rudolf J. Kaaks, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

International Agency for Research
on Cancer
Lyon, France 69008

E-Mail: kaaks@iarc.fr

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

A western lifestyle, characterized by low rates of energy expenditure and a high-energy diet rich in saturated fats and refined carbohydrates, is associated with high incidence of breast cancer in women. Excess energy intake causes elevated blood levels of glucose and insulin, factors that induce fatty acid synthesis and are implicated in the etiology of breast cancer. We recently discovered a highly specific interaction between ACCalpha and the protein coded by the breast cancer susceptibility gene BRCA1, which further supports a possible central role of lipogenic enzymes in breast cancer development.

The above data led us to hypothesize that certain alleles of genes involved in cellular fatty acid synthesis confer increased breast cancer susceptibility. In a series of 46 breast cancer cases, we are systematically searching the coding and regulatory regions of in seven selected candidate genes for polymorphic variations. Using a large case-control study nested within the EPIC cohort, we will examine associations of these sequence variants, and their haplotypes, with breast cancer risk.

The key research accomplishments of the first year of the award are identification of sequence variations and assessement of haplotypes of the *ACCalpha, AMPKalpha1, AMPKalpha2* and *ChREBP* genes coding for key lipogenic enzymes and their principal regulatory factors to be examined for association with breast cancer risk.

14. SUBJECT TERMS
Fatty acid synthesis; Genetic polymorphism; Dietary fats; Excess weight, Breast cancer risk

15. NUMBER OF PAGES
18

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

A western lifestyle, characterized by low rates of energy expenditure and a high-energy diet rich in saturated fats and refined carbohydrates, is associated with high incidence of breast cancer in women. This type of lifestyle induces storage of excess energy in the form of triglycerides, produced either from the diet fatty acids or from those synthesized *de novo*. Excess energy intake and obesity also cause insulin resistance, which is associated with elevated blood levels of glucose and insulin, factors that induce fatty acid synthesis in different tissues and which have been implicated in the etiology of various cancer types including that of the breast [1-12].

Several studies have demonstrated high levels of key fatty acid synthesis enzymes – fatty acid synthase (FAS) and acetyl-CoA carboxylase alpha (ACCalpha) – in human breast cancer as well as in other tumor types [13-16]. FAS inhibitors have been shown to delay tumor progression in xenograft breast cancer models and to induce apoptosis of breast carcinoma cells [17-23]. We recently discovered a highly specific interaction between ACCalpha and the protein coded by the breast cancer susceptibility gene BRCA1 [24], which further supports a possible central role of lipogenic enzymes in breast cancer development.

The above observations lead us to hypothesize that genes involved in cellular fatty acid synthesis may be centrally implicated in mammary gland carcinogenesis and that polymorphic alleles that increase the expression or activity of these genes confer increased breast cancer susceptibility. The specific aims of the proposed study are:

- to search exhaustively for sequence variations in seven selected genes coding for key lipogenic enzymes (*ACCalpha*, *FAS*) and their principal regulatory factors (*AMPKalpha1*, *AMPKalpha2*, *ChREBP*, *SREBP1*, *NFYA*);
- to examine associations of these sequence variations with breast cancer risk, using a large case-control study nested within the *European Prospective Investigation into Cancer and Nutrition* (EPIC) – a prospective cohort in ten Western European countries; and
- to examine interactions between the genetic variants and lifestyle factors such as excess weight and estimated intakes of different types of fats, in determining breast cancer risk.

BODY

The following accomplishments of the tasks of the approved Statement of Work applicable to the first year of the award have been achieved:

Task 1: *Selection of cases and controls, using the established eligibility and matching criteria, and extraction of a database with relevant information from questionnaires and anthropometry (Months 1-4).*

A total of 2510 incident cases of breast cancer were identified, with blood samples taken before cancer diagnosis. To these cases, a total of 3636 control subjects were matched. Matching factors were age at blood donation, EPIC study center of recruitment into the cohort, menopausal status at blood donation, and phase of menstrual cycle (for premenopausal women). By the end of 2005, with a next round of follow-up to identify further incident breast cancer cases, we plan to extend the numbers of cases and controls to about 3000 incident breast cancer cases, and about 4000 matched control subjects.

Task 2: *Retrieval of buffy coat samples from the central EPIC storage facility, and completion of DNA extraction (for ~1000 cases and ~1000 controls for whom DNA has not been extracted yet); preparation of microwell plates with DNA samples, to be ready for PCR (Months 2-12).*

For a total of 1719 cases of breast cancer and 2844 control subjects, DNA was extracted from buffy coat samples, as part of a previous project. For the additional 791 cases currently identified, and their 791 matched controls, all with blood samples stored at the central biorepository at IARC, DNA extraction is currently ongoing.

Task 3: *Exhaustive SNP discovery in all candidate genes by resequencing of DNA (exons and potential regulatory elements) from 46 breast cancer patients (Months 1-24).*

During 2004, we completed the laboratory development steps required to enable high-throughput resequencing at IARC. One important milestone was programming of a Laboratory Information Management System (LIMS) that can track the flow of samples and data in large scale moderately automated projects. The other milestone was development of the automated lab process for resequencing, using dye-primer sequencing chemistry. The dye-primer chemistry has the advantage of giving more even peak heights than dye-terminator chemistry, thus making it more appropriate for heterozygote detection. In addition, the chemistry for dye-primer sequencing is more cost effective than for dye-terminator. We improved upon the publicly available primer selection software Primer3 <http://www-genome.wi.mit.edu/genome_software/other/primer3.html.> by incorporating automatic addition of M13 tails, more robust anti-hairpin protection, and more robust anti-primer dimer protection.

The gene management and resequencing workflow can be viewed as a five step process:

1. *Bio-informatic analysis of gene structure*:
   The genomic structure of the 5' UTR, 3' UTR, and coding regions of the six selected candidate genes (FAS, AMPKalpha1, AMPKalpha2, ChREBP, SREBP1, NFYA) was assessed using data from 3 public databases:
   - the HUGO Gene Nomenclature website <http://www.gene.ucl.ac.uk/nomenclature/>,
   - the UC Santa Cruz genome browser <http://genome.ucsc.edu/>, and
   - PubMed/ Genbank <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>

   (Table 1).
   Genomic and cDNA sequences were downloaded and positions of all of the splice junctions were

confirmed **(Appendix 1)**. We had already screened a 7th gene that is relevant to this study, ACCalpha, for sequence variants in a series of 49 breast cancer familial cases in our previous study (ref 25).

2. *Primer design*:
   Once we are confident of gene structure, M13 tailed primers are designed. We have completed design of primer pairs sufficient to PCR amplify all of the exons of each of the 6 genes that we have proposed to resequence. Primers for 3 of the genes, AMPKalpha1 (PRKAA1), AMPKalpha2 (PRKAA2), and ChREBP (WBSCR14), were ordered and tested in late 2004. Primers for the fourth gene, FAS (which has by far the most exons of the genes that we need to resequence) were ordered in January 2005.

3. *Selection of a DNA sample set fo systematic resequencing*:
   Resequencing is being carried on DNA samples from lymphoblastoid cell lines, established from 46 high-risk breast cancer cases, 1 chimpanzee lymphoblastoid line, and 1 negative control. The breast cancer cases were all chosen from differenty families, and been previously screened for, and found not to carry, clearly deleterious mutations in BRCA1 or BRCA2. All of the cell lines were cultured and DNA preps prepared by September 2004.

   The 6 genes that we are resequencing are comprised of a total of 110 exons. Resequencing of the exons from exon 2 through the 3' UTR does not require any analysis beyond steps 1 and 2 above. However, on the other hand, resequencing of exon 1 and the proximal promoter of each gene benefits from a comparative genomics analysis. Therefore, we describe these two processes separately.

4. *Standard exon resequencing*:
   Our capillary sequencer, a 96-capillary Spectrumedix 9610, was delivered in September 2004. We completed sequencing process development in November 2004. In December we began resequencing 3 of the genes under study, AMPKalpha1 (PRKAA1), AMPKalpha2 (PRKAA2), and ChREBP (WBSCR14). As of the middle of February, we had completed the standard exon resequencing for AMPKalpha1 and AMPKalpha2; 5 standard exons of ChREBP remain to be done. Between these three genes, we have completed the analysis of 29 exons. Thus in ~2.5 months we executed just over 25% of the total resequencing (including the exon 1 and proximal promoters). Our resequencing results are summarized in **Appendices 2.1 to 2.3**.

5. *Sequencing of proximal promoter, and other regulatory elements*:
   Part of our goal is to determine if there are sequence variants in these gene's transcriptional regulatory elements that might alter gene expression. In order to do this, we are taking a comparative genomics approach to identification of the proximal promoter and other potential transcriptional regulatory elements. For each gene of interest, we make a nucleotide multiple sequence alignment covering from ~10,000 bp upstream of exon 1 all the way to exon 2. The alignment includes the human sequence and orthologous sequences from at least 3 different orders of Mammals: rodents (mouse or rat, whichever genomic sequence is more complete across the region of interest), carnivores (dog), artiodactyls (cow) and marsupials (opossum). In fact, the release in January 2005 of apparently high-quality cow and opossum genome sequence assemblies is making our approach more robust than it might otherwise have been. In this approach, potential transcriptional regulatory elements are recognized as non-exon sequences that are conserved across at least 4 of the 5 sequences represented. By this approach, we generally see a proximal promoter that extends ~200 bp upstream (or sometimes downstream) of exon 1 and 3-4 other conserved sequence elements of 50-200 bp that merit resequencing.

At the rate of progress we have been making since the beginning of December 2004, we expect to finish the resequencing phase of this project on schedule, before the end of 2005.

Table 1. *Selected candidate genes*

Gene name	*Gene product*	*Chromosome Band*	*Genomic Size*	*Exons*	*SNP discovery*
ACCalpha	Acetyl-CoA carboxylase alpha	17q12	324,977 bp	60	done
FAS	Fatty acid synthase	17q25.3	19,893 bp	43	to be done
AMPKalpha1	AMP-activated protein kinase alpha 1 catalytic subunit	5p13.1	38,809 bp	11	done
AMPKalpha2	AMP-activated protein kinase alpha 2 catalytic subunit	1p32.2	63,103 bp	9	done
SREBP1	Sterol regulatory element binding protein type 1	17p11.2	24,941 bp	20	to be done
NFYA	CCAAT-binding factor/nuclear factor-Y	6p21.1	26,027 bp	10	to be done
ChREBP	Carbohydrate response element binding protein	7q11.23	31,347bp	17	done

Task 4: *Determination of haplotypes and haplotype-tagging SNPs, using specialized software (Months 3-24).*

Within an exhaustive list of polymorphisms in all coding and regulatory regions of the selected candidate genes it should in principle be possible to identify disease-causing polymorphisms directly, by e.g. multivariate regression modeling. Certain polymorphisms, however, particularly in regulatory sequences, may be missed. A risk-associated haplotype might indicate the presence of a causal, yet to be identified, polymorphism that is in linkage disequilibrium with this haplotype.

The haplotypes of the genes *AMPKalpha1*, *AMPKalpha2* and *ChREBP* – almost fully sequenced in the 46 subjects – have been assessed. The software PHASE [26] is used for estimation of haplotypes and the algorithm described by Stram et al [27] is used for identification of haplotype-tagging SNPs.

In a preliminary study of the *ACCalpha* gene in 453 breast cancer cases and 469 control subjects from France we found significant associations of breast cancer risk with four common *ACCalpha* haplotypes (ref 25). As part of the current project, we typed the same four haplotype-tagging SNPs as in our previous study, for 1719 breast cancer cases and 2844 matched controls within the EPIC study. Results of this first analysis within EPIC, did not show the associations observed in our previous study.

We hypothesized that our initial, positive results could have been due to linkage disequilibrium between the *ACCalpha* assessed risk haplotypes and some yet undetected causal noncoding variants. Using the program "Haploview" [28], we have therefore started examining in greater depth the haplotypes, composed of the SNPs spanning the *ACCalpha* gene region, as documented in the HapMap project (http://www.hapmap.org). The algorithm of Gabriel et al [29] was used for the definition of linkage disequilibrium blocks. Haplotypes were estimated using an accelerated EM algorithm similar to the partition/ligation method described in Qin et al [30]. This analysis led to the identification of 7 additional htSNPs, that are currently (February-March 2005) also being typed for the 1719 breast cancer cases and 2844 control subjects from EPIC. We plan to extend this analysis also to a total of about 3000 breast cancer cases, and over 3000 matched controls, within the EPIC cohorts.

KEY RESEARCH ACCOMPLISHMENTS

The key research accomplishments emanating from the research performed during the first year of the award are:

1. Discovery of sequence variations through the exhaustive resequencing of the coding and regulatory regions of the *AMPKalpha1*, *AMPKalpha2* and *ChREBP* genes coding for principal regulatory factors of key lipogenic enzymes (**Appendix 1**, **Appendices 2.1-2.3**);
2. Assessement of haplotypes and selection of haplotype-tagging SNPs in the *ACCalpha, AMPKalpha1*, *AMPKalpha2* and *ChREBP* genes to be examined for association with breast cancer risk, using a case-control study nested within the EPIC;
3. Selection of a series of over 2500 breast cancer cases and over 3500 control subjects within the EPIC cohorts, for genotyping of haplotype-tagging SNPs in the selected candidate genes, and analysis of gene-breast cancer associations.

REPORTABLE OUTCOMES

None, for year 1.

CONCLUSIONS

Our project is well on schedule, for the identification of SNPs in our candidate genes (systematic resequencing), selection of haplotype-tagging SNPs, selection of case and control subjects for nested case-control analysis within the EPIC cohorts, and extraction of DNA from buffy coat samples of cases and controls.

REFERENCES

1. Kaaks R. Nutrition, hormones, and breast cancer: is insulin the missing link? Cancer Causes Control 1996;7:605-25.
2. Kaaks R, Lukanova A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. Proc Nutr Soc 2001;60:91-106.
3. Kaaks R, Lukanova A. Effects of weight control and physical activity in cancer prevention: role of endogenous hormone metabolism. Ann N Y Acad Sci 2002;963:268-81.
4. Augustin LS, Dal Maso L, La Vecchia C, Parpinel M, Negri E, Vaccarella S, Kendall CW, Jenkins DJ, Francesch S. Dietary glycemic index and glycemic load, and breast cancer risk: a case-control study. Ann Oncol 2001;12:1533-38.
5. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, Rosner B, Speizer FE, Pollak M. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet 1998;351:1393-96.
6. Toniolo P, Bruning PF, Akhmedkhanov A, Bonfrer JM, Koenig KL, Lukanova A, Shore RE, Zeleniuch-Jacquotte A. Serum insulin-like growth factor-I and breast cancer. Int J Cancer 2000;88:828-32.
7. Muti P, Quattrin T, Grant BJ, Krogh V, Micheli A, Schunemann HJ, Ram M, Freudenheim JL, Sieri S, Trevisan M, Berrino F. Fasting glucose is a risk factor for breast cancer: a prospective study. Cancer Epidemiol Biomarkers Prev 2002;11:1361-68.
8. Bianchini F, Kaaks R, Vainio H. Overweight, obesity, and cancer risk. Lancet Oncol 2002;3:565-74.
9. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94:606-16.
10. Li BD, Khosravi MJ, Berkel HJ, Diamandi A, Dayton MA, Smith M, Yu H. Free insulin-like growth factor-I and breast cancer risk. Int J Cancer 2001;91:736-39.
11. Kaaks R, Lundin E, Rinaldi S, Manjer J, Biessy C, Soderberg S, Lenner P, Janzon L, Riboli E, Berglund G, Hallmans G. Prospective study of IGF-I, IGF-binding proteins, and breast cancer risk, in northern and southern Sweden. Cancer Causes Control 2002;13:307-16.
12. Krajcik RA, Borofsky ND, Massardo S, Orentreich N. Insulin-like growth factor I (IGF-I), IGF-binding proteins, and breast cancer. Cancer Epidemiol Biomarkers Prev 2002;11:1566-73.
13. Shurbaji MS, Kalbfleisch JH, Thurmond TS. Immunohistochemical detection of a fatty acid synthase (OA-519) as a predictor of progression of prostate cancer. Hum Pathol 1996;27:917-21.
14. Milgraum LZ, Witters LA, Pasternack GR, Kuhajda FP. Enzymes of the fatty acid synthesis pathway are highly expressed in in situ breast carcinoma. Clin Cancer Res 1997;3:2115-20.
15. Rashid A, Pizer ES, Moga M, Milgraum LZ, Zahurak M, Pasternack GR, Kuhajda FP, Hamilton SR. Elevated expression of fatty acid synthase and fatty acid synthetic activity in colorectal neoplasia. Am J Pathol 1997;150:201-08.
16. Pizer ES, Lax SF, Kuhajda FP, Pasternack GR, Kurman RJ. Fatty acid synthase expression in endometrial carcinoma: correlation with cell proliferation and hormone receptors. Cancer 1998;83:528-37.
17. Kuhajda FP, Jenner K, Wood FD, Hennigar RA, Jacobs LB, Dick JD, Pasternack GR. Fatty acid synthesis: a potential selective target for antineoplastic therapy. Proc Natl Acad Sci U S A 1994;91:6379-83.
18. Pizer ES, Jackisch C, Wood FD, Pasternack GR, Davidson NE, Kuhajda FP. Inhibition of fatty acid synthesis induces programmed cell death in human breast cancer cells. Cancer Res 1996;56:2745-47.
19. Pizer ES, Wood FD, Heine HS, Romantsev FE, Pasternack GR, Kuhajda FP. Inhibition of fatty acid synthesis delays disease progression in a xenograft model of ovarian cancer. Cancer Res 1996;56:1189-93.
20. Pizer ES, Thupari J, Han WF, Pinn ML, Chrest FJ, Frehywot GL, Townsend CA, Kuhajda FP. Malonyl-coenzyme-A is a potential mediator of cytotoxicity induced by fatty-acid synthase inhibition in human breast cancer cells and xenografts. Cancer Res 2000;60:213-18.

21. Kuhajda FP, Pizer ES, Li JN, Mani NS, Frehywot GL, Townsend CA. Synthesis and antitumor activity of an inhibitor of fatty acid synthase. Proc Natl Acad Sci U S A 2000;97:3450-54.

22. Thupari JN, Pinn ML, Kuhajda FP. Fatty acid synthase inhibition in human breast cancer cells leads to malonyl-CoA-induced inhibition of fatty acid oxidation and cytotoxicity. Biochem Biophys Res Commun 2001;285:217-23.

23. Alo PL, Visca P, Trombetta G, Mangoni A, Lenti L, Monaco S, Botti C, Serpieri DE, Di Tondo U. Fatty acid synthase (FAS) predictive strength in poorly differentiated early breast carcinomas. Tumori 1999;85:35-40.

24. Magnard C, Bachelier R, Vincent A, Jaquinod M, Kieffer S, Lenoir GM, Venezia ND. BRCA1 interacts with acetyl-CoA carboxylase through its tandem of BRCT domains. Oncogene 2002;21:6729-39.

25. Sinilnikova OM, Ginolhac SM, Magnard C, Leone M, Anczukow O, Hughes D, Moreau K, Thompson D, Coutanson C, Hall J, Romestaing P, Gerard JP, Bonadona V, Lasset C, Goldgar DE, Joulin V, Venezia ND, Lenoir GM. Acetyl-CoA carboxylase alpha gene and breast cancer susceptibility. Carcinogenesis. 2004;12:2417-24.

26. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978-89.

27. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC. Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum Hered. 2003;55:27-36.

28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263-5.

29. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002;296:2225-9.

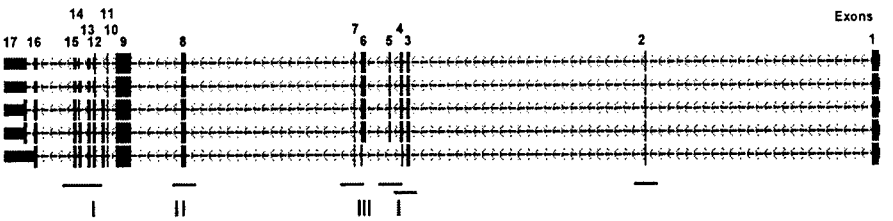
30. Qin ZS, Niu T, Liu JS. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. Am J Hum Genet. 2002;71:1242-7.

APPENDICES

Appendix 1: Genomic structure of the candidate genes examined

Slide 1

CHREBP/WBSCR14 gene

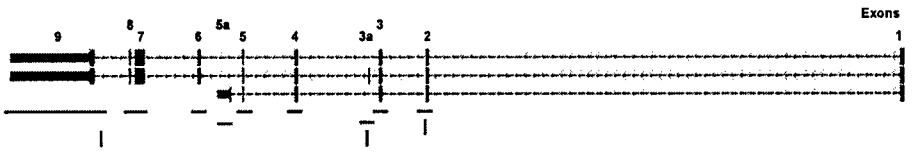


Resequenced regions
— done
I SNPs identified

Position: chr7:72,452,174-72,483,521
Chromosome Band: 7q11.23
Genomic Size: 31,347bp
Exons: 17

Slide 2

AMPKalpha1/PRKAA1 gene

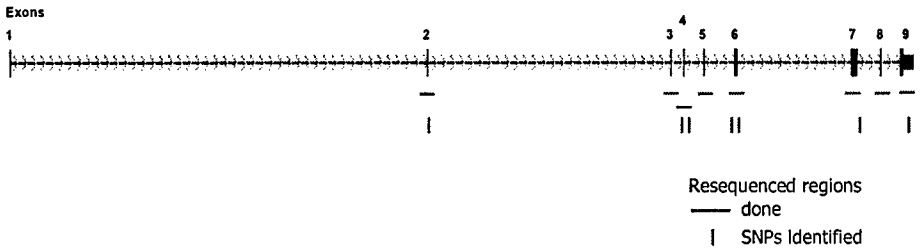


Resequenced regions
— done
I SNPs identified

Position: chr5:40,795,238-40,834,046
Chromosome Band: 5p13.1
Genomic Size: 38,809 bp
Exons: 11

Slide 3

AMPKalpha2/PRKAA2 gene



Resequenced regions
— done
| SNPs Identified

Position: chr1:56,823,040-56,886,142
Chromosome Band: 1p32.2
Genomic Size: 63,103 bp
Exons: 9

Slide 4

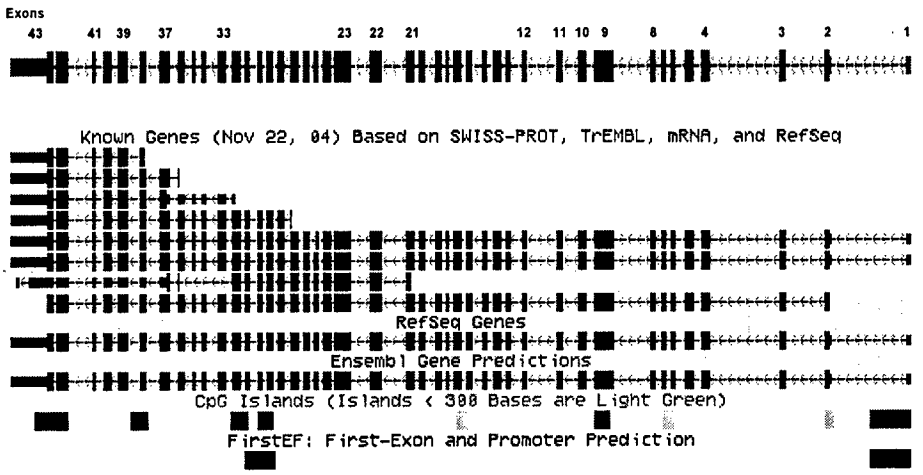
FASN gene

Resequenced regions
— done
| SNPs identified

Position: chr17:77,629,503-77,649,395
Chromosome Band: 17q25.3
Genomic Size: 19,893 bp
Exons: 43

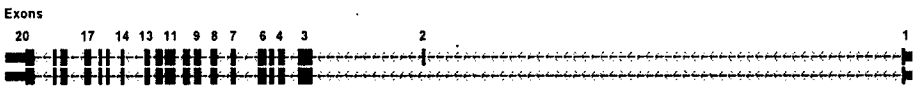
Slide 5

FASN gene: multiple alternative transcripts



Slide 6

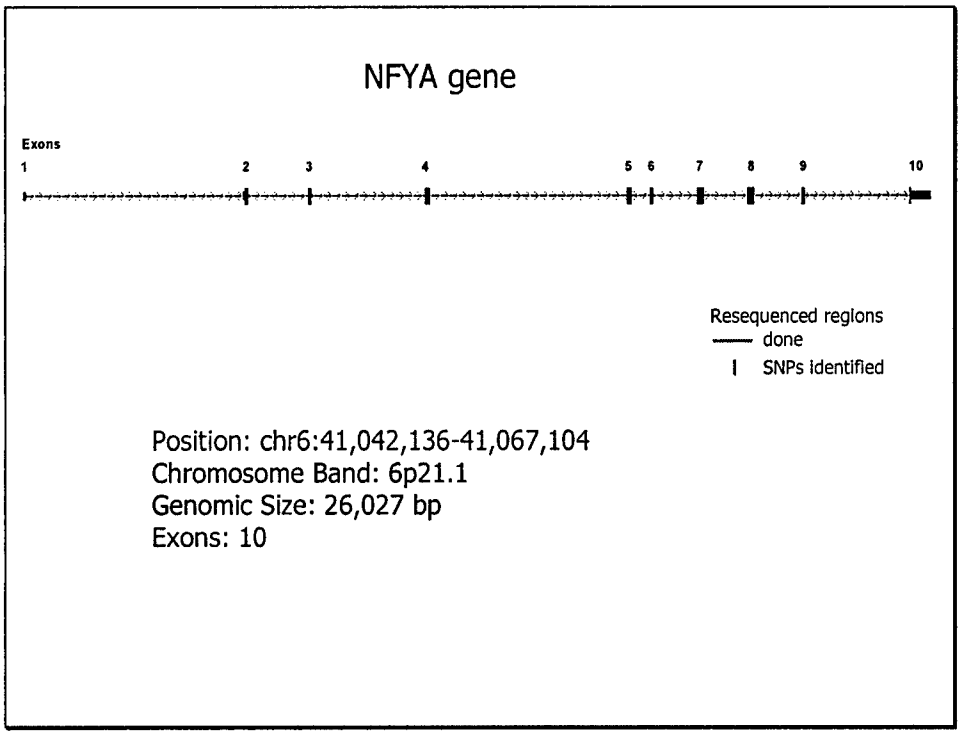
SREBP1 gene



Resequenced regions
—— done
| SNPs identified

Position: chr17:17,656,110-17,681,050
Chromosome Band: 17p11.2
Genomic Size: 24,941 bp
Exons: 20

Slide 7


NFYA gene
Exons
1
2
3
4
5
6
7
8
9
10
Resequenced regions
done
SNPs identified
Position: chr6:41,042,136-41,067,104
Chromosome Band: 6p21.1
Genomic Size: 26,027 bp
Exons: 10

Appendix 2.1

AMPKalpha1/PRKAA1: Summary resequencing results.

PRKAA1 is a gene with 9 commonly used exons and 2 alternative exons. To date, we have resequenced 6 of these.
Resequencing of exons 1, 7, 8, and the proximal promoter is in progress

Sample ##	ID	Exon 2 amp BQ	Exon 3 amp BQ	Exon 3A amp BQ	Exon 4 amp BQ	Exon 5 amp BQ	Exon 5A amp BQ	amp CR	Exon 6 amp BQ	Exon 7 amp BQ	Exon 8 amp BQ	Exon 9 amp BQ	amp CR
mutscreen name		229 ins TA		192 C>G								49 G>A	
exon name		i+15 ins TA		i+79 C>G								i-42 G>A	
note		intron		intron								intron	
layout position		25,756		28,337								40,122	
SNP surrounding sequence		TTGGCAAAC		actaaagatctag								##########	
SNP mutscreen name		229 ins TA		192 C>G								49 G>A	
SNP gene location		Intron 2		Intron 3a								Intron 8	
SNP description		IVS2+15insTA		IVS3a+79C>G								IVS8-42G>A	
Position:Hum Genome-May200		Chr5:40813289		Chr5:40810708								Chr5:40798923	
rs#		unknown		unknown								unknown	
Freq minor allele our sample		0.250		0.033								0.011	
1	IARC-1479	TT	**	CC	**	**	--	--	**	**	**	GG	**
2	IARC-1507	TT	**	CC	**	**	**	--	**	--	--	--	**
3	IARC-1526	TT	**	CC	**	**	**	**	**	**	**	GG	**
4	IARC-1568	TT	**	CG	**	--	**	**	**	**	**	GG	**
5	IARC-1679	T ins	**	CC	**	**	**	**	**	**	**	GG	**
6	IARC-1789	TT	**	CC	**	**	**	**	**	**	**	GG	**
7	IARC-1802	T ins	**	CC	**	**	**	**	**	**	**	GG	**
8	IARC-1807	ins ins	**	CC	**	**	**	**	--	**	**	GG	**
9	IARC-1909	TT	**	CC	**	**	**	**	**	**	**	GG	**
10	IARC-1928	TT	**	CC	**	**	**	**	**	**	--	--	**
11	IARC-1967	T ins	**	CC	**	**	**	**	**	**	**	GG	**
12	IARC-1981	T ins	**	CC	**	**	**	**	**	**	**	GG	--
13	IARC-2026	T ins	**	CC	**	**	**	**	**	**	**	GG	**
14	IARC-2117	TT	**	CG	**	**	**	**	**	**	**	GG	**
15	IARC-2166	T ins	**	CC	**	**	**	**	**	**	**	GG	**
16	IARC-2167	T ins	**	CC	**	**	**	**	**	**	**	GG	**
17	IARC-2181	TT	**	CC	**	**	**	**	**	**	**	GG	**
18	IARC-2193	T ins	**	CC	**	**	**	**	**	**	**	GG	**
19	IARC-2212	T ins	**	CC	--	**	**	**	**	**	--	GG	**
20	IARC-2214	ins ins	**	CC	**	**	**	**	**	**	**	GG	**
21	IARC-2247	T ins	**	CC	**	**	**	**	**	**	**	GG	**
22	IARC-2351	TT	**	CC	**	**	**	**	--	**	**	GG	**
23	IARC-2394	TT	**	CC	**	**	**	**	**	**	**	GG	**
24	IARC-2443	T ins	**	CC	**	**	**	**	**	**	**	GG	**
25	IARC-2461c	TT	**	CC	**	**	**	**	**	**	**	GG	**
26	IARC-2472	TT	**	CC	**	**	**	**	**	**	**	GG	**
27	IARC-2477	TT	**	CC	**	**	**	**	**	**	**	GG	**
28	IARC-2525	TT	**	CC	**	**	**	**	**	**	**	GG	**
29	IARC-2526	T ins	**	CC	**	**	**	**	**	**	**	GG	**
30	IARC-2527	TT	**	CC	**	**	**	**	**	**	**	GG	**
31	IARC-2528	TT	**	CC	**	**	**	**	**	**	**	GG	**
32	IARC-2529	TT	**	CC	**	**	**	**	**	**	**	GG	**
33	IARC-2530	T ins	**	CC	**	**	**	**	**	**	**	GG	**
34	IARC-2531	TT	**	CC	**	**	**	--	**	**	**	GG	**
35	IARC-2533	TT	**	CC	**	**	**	**	**	**	**	GG	**
36	IARC-2534	TT	**	CG	**	**	**	**	**	**	**	GG	**
37	IARC-2535	ins ins	**	CC	**	**	**	**	**	**	**	GG	**
38	IARC-2536	ins ins	**	CC	**	**	**	**	**	--	**	GG	**
39	IARC-2537	TT	**	CC	**	**	**	**	**	**	**	AA	**
40	IARC-2538	TT	**	CC	**	--	**	**	**	--	**	GG	**
41	IARC-2539	TT	**	CC	--	**	--	--	**	**	**	GG	**
42	IARC-2540	T ins	**	CC	**	**	**	**	**	**	**	GG	**
43	IARC-2541	TT	**	CC	**	**	--	--	**	**	**	GG	**
44	IARC-2542	T ins	**	CC	**	**	**	**	**	**	**	GG	**
45	IARC-2543	TT	**	CC	**	**	**	**	**	**	**	GG	**
46	IARC-2544	TT	**	CC	**	**	**	**	**	**	**	GG	**
47	Chimpanzee	TT	**	CC	**	**	**	**	**	**	**	GG	**
48	Negative	--	--	--	--	--	--	--	--	--	--	--	--
Additional in chimp:		0	0	0	0	0	1	5	0	0	0	0	0

-- resequencing failed
** no sequence variants observed in this amplicon

Sequence variants marked in green were observed once only

Haplotype	i+15 ins TA	freq
1	T	0.717
2	ins	0.250
3	T	0.033

Appendix 2.2

AMPKalpha2/PRKAA2: Summary resequencing results.

PRKAA2 is a gene of 9 exons. Resequencing of 8 of its exons (exons 2 - 9) is complete
Resequencing of exon 1 and the proximal promoter is in progress

Sample		Exon 2	Exon 3	Exon 4		Exon 5	Exon 6		Exon 7		Exon 8	Exon 9	
##	ID	amp BQ	amp BQ	amp BQ		amp BQ	amp BQ		amp BQ	amp CR	amp BQ	amp BQ	amp CR
mutscreen name		100 A>G		90 G>A	189 C>T		59 G>A	166 G>C		342 T>C			186 A>G
exon name		17 A>G		21 G>A	120 C>T		i-59 G>A	49 G>C		i+81 T>C			444 A>G
note		silent		silent	silent		intron	L->F		known!			known!
layout position		34,077		52,058	52,157		55,556	55,663		64,236			67,598
SNP surrounding sequence		##########		##########	##########		##########	##########		GAATCCGAA			##########
SNP mutscreen name		100 A>G		90 G>A	189 C>T		59 G>A	166 G>C		342 T>C			186 A>G
SNP gene location		Exon 2		Exon 4	Exon 4		Intron 5	Exon 6		Intron 7			Exon 9
SNP description		A>G L37L		G>A R117R	C>T H150H		IVS5-59G>A	G>C L204F		IVS7+81 T>C			3'UTR+205A>G
Position:Hum Genome-May2004		Chr1:56852092		Chr1:5687007	Chr1:56870172		Chr1:5687357	Chr1:56873678		Chr1:56882250			Chr1:56885612
rs#		unknown		unknown	unknown		unknown	unknown		rs932447			rs3736568
Freq minor allele:our sample		0.011		0.045	0.011		0.012	0.012		0.398			0.024
1	IARC-1479	AA	**	GA	GT	--	--	--	**	TT	**	**	AA
2	IARC-1507	AA	**	GG	CC	**	GG	GG	--	TC	**	**	AA
3	IARC-1526	AA	**	GG	CC	**	GG	GG	**	TC	**	--	AA
4	IARC-1568	AA	**	GG	CC	**	--	--	**	TC	**	--	AA
5	IARC-1679	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
6	IARC-1789	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
7	IARC-1802	AA	**	GG	CC	**	GG	GG	**	CC	**	**	AA
8	IARC-1807	AA	**	GG	CC	**	GG	GG	**	CC	**	--	--
9	IARC-1909	AA	**	GA	CC	**	--	--	**	TT	**	**	AG
10	IARC-1928	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
11	IARC-1967	AA	**	GA	CC	--	GG	GG	**	TC	**	**	AA
12	IARC-1981	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
13	IARC-2026	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
14	IARC-2117	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
15	IARC-2166	--	**	GG	CC	**	GG	GG	**	CC	**	**	AA
16	IARC-2167	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
17	IARC-2181	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
18	IARC-2193	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
19	IARC-2212	AA	**	--	--	**	GG	GG	**	TC	**	**	AA
20	IARC-2214	AA	**	GG	CC	**	GG	GG	**	CC	**	**	AA
21	IARC-2247	AA	**	GG	CC	**	GG	AG	**	TC	**	**	AA
22	IARC-2351	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
23	IARC-2394	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
24	IARC-2443	AA	**	GG	CC	--	GG	GG	**	TT	**	**	AA
25	IARC-2461c	AA	**	GG	CC	**	--	--	**	TT	**	--	--
26	IARC-2472	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
27	IARC-2477	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
28	IARC-2525	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
29	IARC-2526	AA	**	GG	CC	**	--	--	**	--	**	**	AA
30	IARC-2527	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
31	IARC-2528	--	**	GG	CC	**	GG	GG	**	TT	**	**	AA
32	IARC-2529	AA	**	--	--	**	GG	GG	**	TC	**	**	AA
33	IARC-2530	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
34	IARC-2531	AA	**	GG	CC	--	GG	GG	**	TT	**	**	--
35	IARC-2533	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
36	IARC-2534	AG	**	GG	CC	**	GG	GG	**	CC	**	**	AA
37	IARC-2535	AA	**	GG	CC	**	GG	GG	**	CC	**	**	AA
38	IARC-2536	AA	**	GG	CC	**	AG	GG	**	TT	**	**	AA
39	IARC-2537	AA	**	GG	CC	**	GG	GG	**	TC	**	**	AA
40	IARC-2538	AA	**	GG	CC	--	GG	GG	**	CC	**	**	--
41	IARC-2539	AA	**	GG	CC	**	GG	GG	**	--	**	**	--
42	IARC-2540	AA	**	GG	CC	--	GG	GG	**	TC	**	**	AA
43	IARC-2541	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
44	IARC-2542	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
45	IARC-2543	AA	**	GA	CC	**	GG	GG	**	TC	**	**	AG
46	IARC-2544	AA	**	GG	CC	**	GG	GG	**	TT	**	**	AA
47	Chimpanzee	AA	**	GG	CC	--	GG	GG	**	TT	**	**	AA
48	Negative	--	--	--	--	--	--	--	--	--	--	--	--
Additional in chimp		2	0	5		0	1		4		0	2	

-- resequencing failed
** no sequence variants observed in this amplicon

Sequence variants marked in green were observed once only

The two uncommon variants marked in blue are probably in disequilibrium

Haplotype	Exon 4 90 G>A	Exon 7 342 T>C	Exon 9 186 A>G	Approx freq
1	G	T	A	0.576
2	A	T	A	0.022
3	A	T	G	0.022
4	G	C	A	0.380

Appendix 2.3

CHREBP/WBSCR14: Summary resequencing results.

WBSCR14 is a gene of 15 exons. Resequencing of 11 of its exons (exons 2 - 8, 12-15) is complete.
Resequencing of exon 1 and the proximal promoter, as well as exons 9-11, are in progress.

Sample		Exon 2	Exon 3	Exon 4	Exon 5	Exon 6			Exon 7	Exon 8		Exon 9	Exon 10	Exon 11	Exon 12	Exon 13	Exon 14	Exon 15
##	ID	amp BQ	amp BQ	amp BQ	amp BQ	amp BQ			amp BQ	amp BQ		amp BQ	amp BQ	amp BQ	amp BQ	amp BQ	amp BQ	amp BQ
mutscreen name				208 G>C		73 G>A	175 G>C	211 G>A		204 G>A	263 A>G				136 G>A			
exon name				i+8 G>C		3 G>A	105 G>C	141 G>A		125 G>A	i+14				56 G>A			
note				intron		silent	Q->H	silent		silent	intron				silent			
layout position				22,218		23,432	23,534	23,570		29,970	30,029				33,117			
SNP surrounding sequence				##########		##########	##########	##########		##########	##########				##########			
SNP mutscreen name				208 G>C		73 G>A	175 G>C	211 G>A		204 G>A	263 A>G				136 G>A			
SNP gene location				Intron 4		Exon 6	Exon 6	Exon 6		Exon 8	Intron 8				Exon 12			
SNP description				IVS4+8 G>C		G>A A207A	G>C Q241H	A>G S2533		G>A P342P	IVS8+14 A>G				G>A L626L or X676X in alternative transcript			
Position:Hum Genome- May2004				Chr7:72466304		Chr7:72465	Chr7:7246459	Chr7:7246496		Chr7:7245855	Chr7:7245849				Chr7:72456405			
rs#				unknown		rs12539160	rs3812316	rs799157		rs13236643	unknown				unknown			
Freq minor allele our sample				0,012		0,012	0,113	0,012		0,113	0,012				0,044			
1	IARC-1479	**	**	GG	--	GG	GC	GG	**	GA	AA				GG	**	--	**
2	IARC-1507	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
3	IARC-1526	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
4	IARC-1568	**	**	AG	**	GG	--	GG	**	GA	AA				GG	**	**	**
5	IARC-1679	**	**	--	**	GG	GC	GG	**	GA	AA				GA	**	**	**
6	IARC-1789	**	**	GG	**	GG	GC	GG	**	GA	--				GG	**	**	**
7	IARC-1802	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
8	IARC-1807	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
9	IARC-1909	**	**	GG	--	GG	GC	GG	**	GA	AA				GG	**	**	**
10	IARC-1928	**	**	--	**	GG	GG	GG	**	GG	AA				GG	**	**	**
11	IARC-1967	**	**	GG	**	--	--	--	**	GA	AG				GG	**	**	**
12	IARC-1981	**	**	GG	--	GG	GG	GG	**	GG	AA				GG	**	**	**
13	IARC-2026	**	**	GG	**	GG	GC	GG	**	GA	AA				GG	**	**	**
14	IARC-2117	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
15	IARC-2166	**	**	GG	**	GG	GG	GG	**	AA	AA				GG	**	**	**
16	IARC-2167	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
17	IARC-2181	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
18	IARC-2193	**	**	GG	--	GG	GC	GG	**	GA	AA				GG	**	**	**
19	IARC-2212	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
20	IARC-2214	**	**	GG	--	--	--	--	**	GG	AA				GG	**	**	**
21	IARC-2247	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
22	IARC-2351	**	**	GG	**	GG	GG	GG	**	GG	AA				GA	**	**	**
23	IARC-2394	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
24	IARC-2443	**	**	GG	--	GG	GG	GG	**	GG	AA				GG	**	**	**
25	IARC-2461c	**	**	--	--	--	--	--	**	GG	AA				GG	**	**	**
26	IARC-2472	**	--	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
27	IARC-2477	**	--	GG	**	GG	GC	GG	**	GA	AA				GG	**	**	**
28	IARC-2525	**	**	GG	**	GG	GG	AG	**	GG	AA				GG	**	**	**
29	IARC-2526	**	**	GG	**	GG	GG	GG	**	GG	AA				GA	**	**	**
30	IARC-2527	**	**	GG	**	GG	GG	GG	**	GG	AA				GA	**	**	**
31	IARC-2528	**	--	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
32	IARC-2529	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
33	IARC-2530	**	**	GG	--	AG	GG	GG	**	GG	AA				GG	**	**	**
34	IARC-2531	**	**	GG	--	GG	GG	GG	**	GG	AA				GG	**	**	**
35	IARC-2533	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
36	IARC-2534	**	**	GG	**	--	--	--	**	--	--				--	**	**	**
37	IARC-2535	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
38	IARC-2536	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
39	IARC-2537	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
40	IARC-2538	**	**	--	--	GG	GG	GG	**	--	--				GG	**	**	**
41	IARC-2539	**	--	GG	**	--	--	--	**	--	--				GG	**	**	**
42	IARC-2540	**	**	GG	--	GG	GG	GG	**	GG	AA				GG	**	**	--
43	IARC-2541	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	--
44	IARC-2542	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
45	IARC-2543	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	**
46	IARC-2544	**	**	GG	**	GG	GG	GG	**	GG	AA				GG	**	**	--
47	Chimpanzee	**	**	--	--	GG	GG	GG	**	--	--				GG	**	**	--
48	Negative	--	--	--	--	--	--	--	--	--	--				--	--	--	--
Additional in chimp		0	0	0	0	1			1	0					1	1	3	0

-- resequencing failed
** no sequence variants observed in this amplicon

Sequence variants marked in green were observed once only

The two uncommon variants marked in blue are in disequilibrium

Haplotype	Exon 6 105 G>C	Exon 8 125 G>A	Exon 12 56 G>A	Approx freq
1	G	G	G	0.830
2	G	G	A	0.044
3	C	A	G	0.125